Journal of Computers Vol. 30 No. 6, 2019, pp. 216-222
doi:10.3966/199115992019123006017

A Bread Recognition System Based on Faster R-CNN

WenBin Liu, Jia Guo, HeZhi Lin*, LianFen Huang, ZhiBin Gao

Department of Communication Engineering, Xiamen University, Xiamen, China
linhezhi@xmu.edu.cn

Received 13 January 2019; Revised 17 January 2019; Accepted 12 March 2019

Abstract. Nowadays, with the advancement of target recognition technology, many scenes have been semi-automatic or automated, and automated retailing has naturally received much attention. In this paper, the deep learning method is applied to target recognition, and an accurate bread recognition system based on Faster R-CNN deep learning network is built. This paper introduces the hardware structure and working principle of the system, and describes in detail how to implement this bread identification system, and finally tests the speed and accuracy of the system through experiments. The experimental results show that the deep learning method based on Faster R-CNN can accurately identify the bread and obtain the variety of bread. Because the system has the advantages of fast detection speed and high recognition accuracy, the system can meet the market demand and has good promotion.

Keywords: bread recognition system, deep learning, Faster R-CNN, identification

1 Introduction

1.1 Motivation

In recent years, under the background of the rapid development of mobile Internet and artificial intelligence, there have been unmanned restaurants and unmanned bakeries. Because of its convenience, flexibility and many other advantages, it is very popular. The efficiency and accuracy of self-checkout directly affects the satisfaction of customers in unmanned stores [5].

1.2 Current Technology

There are three main methods of checkout: one is traditional manual settlement. In this way, low efficiency, high cost, and fatigue caused by long hours of work may cause many errors to affect the customer's experience. Another method is bar code technology, which is used to supplement the barcode identification of each product to assist the identification. Although the efficiency is greatly improved compared with the traditional manual settlement, the bar code technology often requires manual because of the uncertainty of the bar code printing position. Auxiliary, the degree of automation is not high. The third way is RFID (Radio Frequency Identification) technology, similar to bar codes, which assigns a unique number to each item and identifies the item by wireless signal communication. Although this method is highly automated, wireless signals are still widely used because they are susceptible to interference, high cost, and still not suitable for unpackaged bread and other products. With the substantial improvement of computer computing power and the cost of image acquisition equipment, the product identification technology based on target recognition has received more and more attention due to its low cost, high degree of automation and wide application range [5-6].

In order to improve the recognition accuracy and speed, the model is continuously optimized, and the Fast R-CNN and Faster R-CNN models are generated. The Faster R-CNN model has great advantages in the speed and accuracy of recognition, and it has been applied in many recognition research such as vehicle identification and its effect is outstanding. This paper proposes an accurate bread identification

* Corresponding Author

system based on Faster R-CNN for the unmanned bread shop. The system has low cost, high precision, good real-time performance and good generalization [5-6].

1.3 Context Outlines

The second chapter of this paper describes the hardware structure of the system, the establishment of the data set and how to train and generate the model of bread recognition in detail.

The third chapter of this paper describes the results obtained from the experiments. The final section explains the conclusions reached.

2 System Design

The goal of the bread identification system based on the Faster R-CNN deep learning network described herein is to obtain the categories and quantities of all the targets in the image to obtain the price of the bread purchased by the consumer. This chapter describes how to implement this bread identification system from both the hardware and software aspects of the system.

2.1 Hardware Design

The hardware structure of the bread identification system described herein consists of an acquisition system and a processing and display system. The acquisition system consists of a bracket, a platform, a camera, and a fill light; the processing terminal and the display form a processing and display system. The bracket provides the camera with a suitable angle to take pictures of the bread on the platform, and the fill light provides suitable light for image acquisition. The processing terminal, as the core of the system, obtains the category of the bread on the platform after processing the captured image.

2.2 Faster R-CNN

Faster R-CNN is an improvement from Fast R-CNN. The Faster R-CNN algorithm first sends the input picture to the conv layers for processing to obtain its feature maps for subsequent RPN layers and fully connected layers. In this paper, VGG16 is used as the conv layers. The RPN layer is a key step in Faster R-CNN. The input and output of the RPN layer are respectively the convolution feature map and the images generated by the image. In the RPN layer, the input convolution feature map is first sent to the $3 \times 3 \times 512$ convolution layer processing. After that, the result is sent to two parallel 1×1 convolution layer, which are the classification layer and the regression layer, respectively. In the classification layer, the anchors of each pixel are subjected to two classification processing to determine whether it is a background or a target; in the regression layer, four coordinate information of the anchors corresponding to each pixel are obtained. The RPN layer is shown in Fig. 1. Next, the proposal generated by the RPN layer and the feature maps generated by the conv layers are sent to the ROI Pooling layer, which uses each proposal to crop the feature maps to obtain a fixed-size proposal feature map. It is then entered into the fully connected layer for target recognition and location tasks. The system block diagram of Faster R-CNN is shown in Fig. 2 [1, 4].

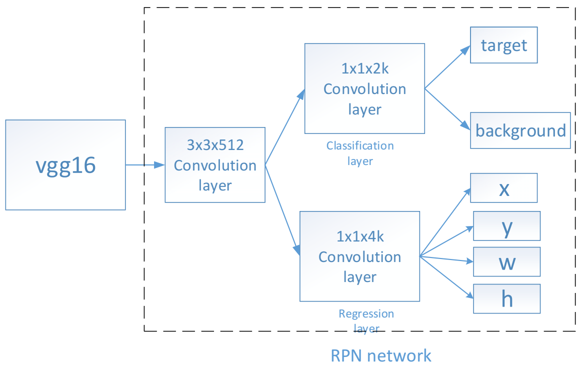


Fig. 1. RPN network

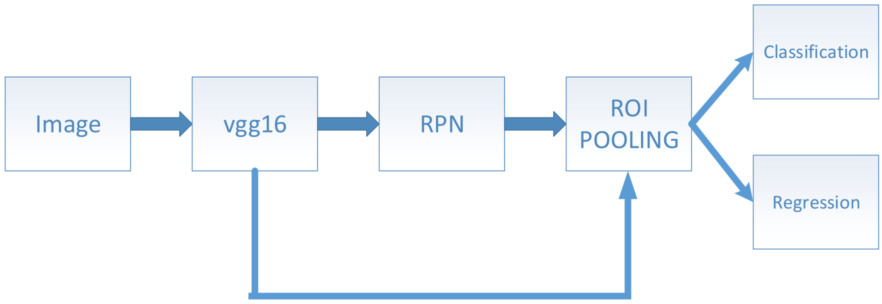


Fig. 2. System diagram of Faster R-CNN

This paper adopts the end-to-end training method. Compared with distributed training, this training method can improve the training speed and reduce the memory used under the premise of ensuring high recognition rate. Faster R-CNN training is based on the already trained vgg16 pre-training model. There are four loss functions for the Faster R-CNN deep learning network: RPN classification, RPN regression, Fast R-CNN classification, and Fast R-CNN regression.

2.3 Prepare the Data Set

In this bread recognition system, in order to simulate the scene of the unmanned bread shop, the research team used the camera to take 507 pictures, and the types and positions of the bread in the photos were randomly placed. Because the pixel of the original image is too high, it will increase the training time and complexity. Therefore, we first preprocess the captured image and unify the original image size to 640×480. Then, for the convenience of training, press 1 to 507 to name the picture accordingly. The original image is then calibrated by the LableImg tool, and an .xml file conforming to the pascal VOC format is generated for training. The labeling process is shown in Fig. 3. There are 13 types of test breads, such as oval bread, cutting bread, walnut flower loaf, cup cake, sesame bread, caterpillar bread, flower bread, smiley bread, sesame flower bread, pineapple bread, blue macaron, grey macaron, yellow macaron. After the calibration is completed, the original image and the calibrated .xml file are respectively placed in the voc2007 directory, and the txt file index required for training and testing is generated, and the image is acquired according to the index during training. The validation set used in this experiment is 55 images randomly extracted from 507 original images. Fig. 4 shows a portion of a randomly extracted validation set image.

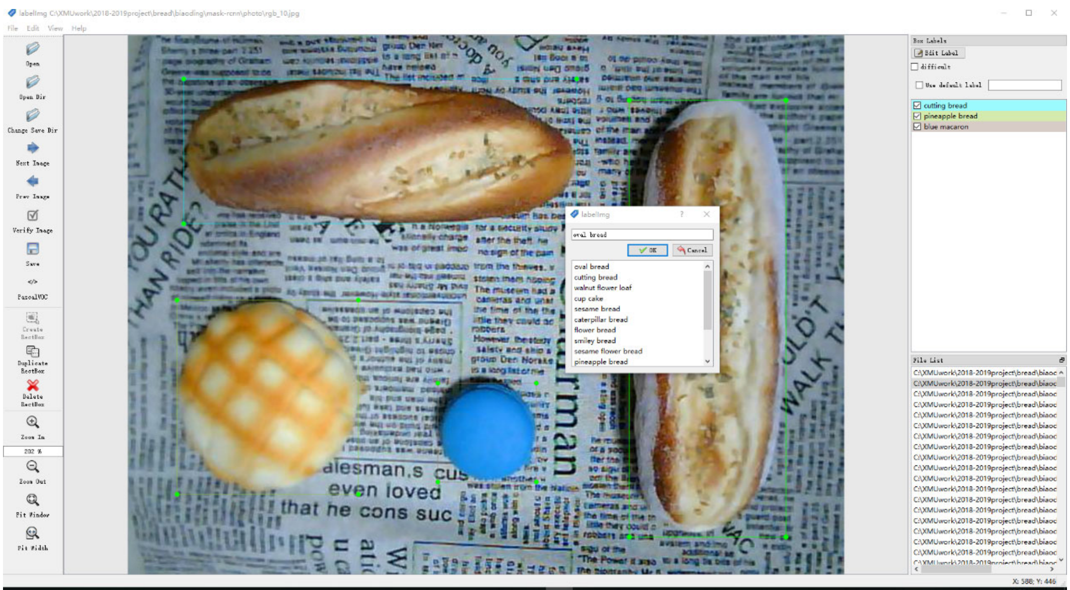
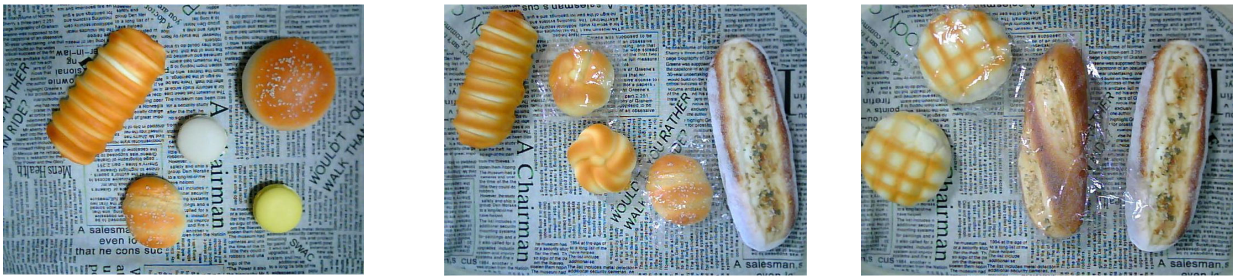


Fig. 3. Label data set



Random picture 1 Random picture 2 Random picture 3

Fig. 4. Partial verification set image

2.4 Train the Faster R-CNN Network

The training environment of the Faster R-CNN deep learning network in this paper is based on tensorflow1.9.0 of python3.5 version under Ubuntu16.04 system, and accelerates training under Nvidia GTX 1080 GPU. The initial training parameters are the parameters of the trained VGG16 model under ImageNet.

Before the training starts, you need to set the total number of all categories and categories. The test described in this article has 13 types of bread. Because of the background, the total number of the types is 14. After that, the pre-training model used for this training is VGG16. The number of iterations of training is 70,000. After the setup is complete, the training begins. The training takes 8 hours, 45 minutes, and 32 seconds. The variation of the loss function with the number of iterations during training is shown in Fig. 5.

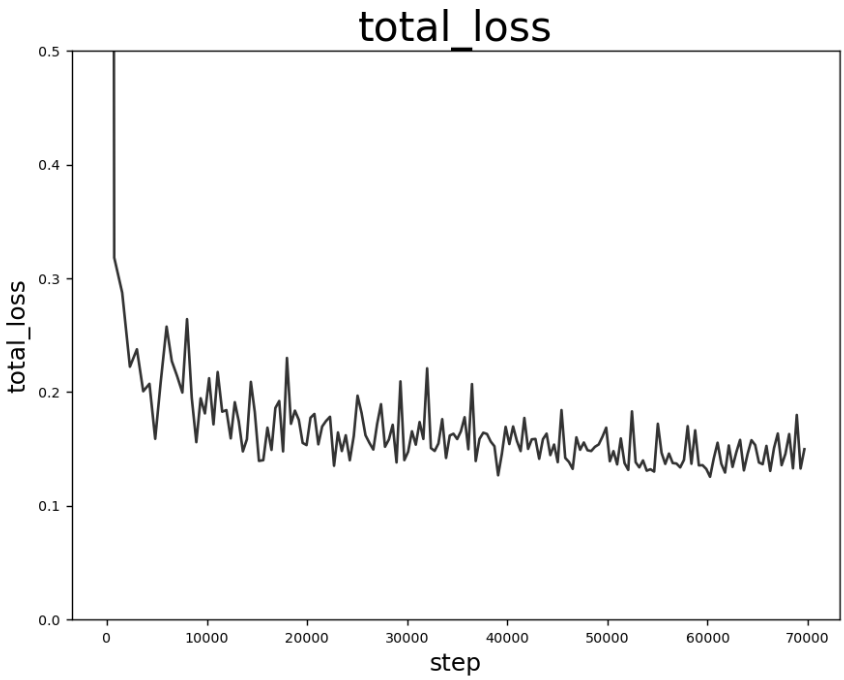
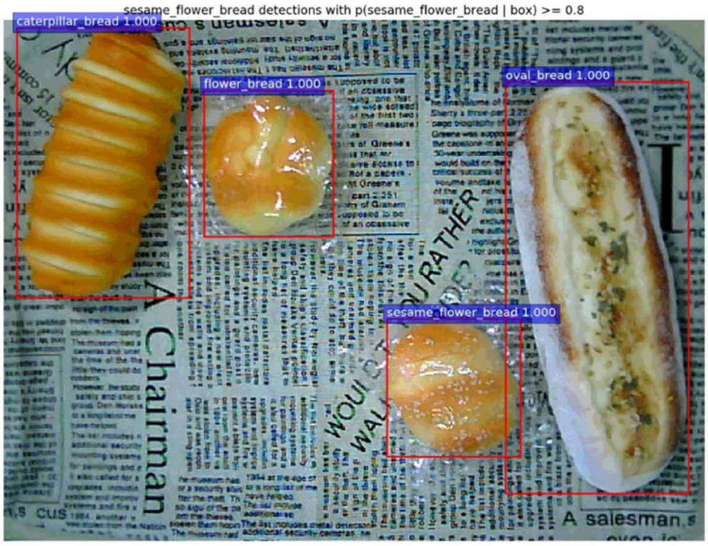


Fig. 5. Change process of loss function

Write code to mark the bread in the input image with a rectangular frame in the output image and display the type of bread. Finally, the running speed of the single image and the type and quantity of the bread in the input image are output for subsequent valuation. Test results shows the system found there was 1 caterpillar bread, 1 flower bread, 1 oval bread and 1 sesame flower bread in the input image. The result is shown in Fig. 6. According to the test results in the verification set, the mAP of the model reaches 0.9914, and the results on the verification set are shown in Table. 1.



Output image

Fig. 6. Result

Table 1. AP value of various types of bread in the verification set

Number	Name	AP
1	Oval bread	1.0000
2	Cutting bread	1.0000
3	Walnut flower bread	1.0000
4	Cup cake	1.0000
5	Sesame bread	0.9679
6	Caterpillar bread	1.0000
7	Flower bread	1.0000
8	Smiley bread	1.0000
9	Sesame flower bread	1.0000
10	Pineapple bread	0.9208
11	Blue macaron	1.0000
12	Grey macaron	1.0000
13	Yellow macaron	1.0000

3 Experimental Result

This chapter tests the performance of the trained model and finds that the bread recognition system has a fast recognition speed, high recognition accuracy and good effect. Compared with the model trained by the SSD model, it has a great advantage in accuracy.

3.1 Test Results

In the training described in Chapter 3, the test set was not added for training. Therefore, in order to verify the accuracy of the bread-based recognition of the model obtained by the training, the research team additionally took 3 randomly placed bread pictures for testing. The pictures of the new test set have changed in height and angle compared to the training set. Some of the results are shown in Fig.7 and Table 2.

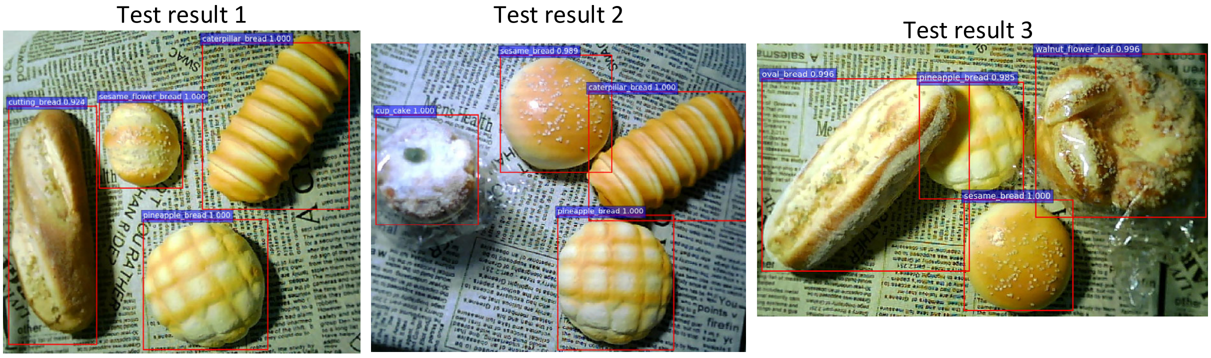


Fig. 7. Test result

Table 2. AP value of various types of bread in the verification set

Number	Test time(s)	Accuracy
1	1.465	100%
2	0.076	100%
3	0.079	100%

It can be seen from the results that the bread identification system can accurately identify the bread and its type in the test set with different light conditions and different shooting angles, and can accurately calculate the number of each type of bread. The accuracy of recognition is close to 100%. The experimental results show that the bread identification system has good robustness and can maintain good accuracy under the conditions of ambient brightness and shooting angle. In the recognition speed, the speed of the first picture detected by the opening is 1.4s, and the recognition speed of the single picture is less than 0.1s, which has good real-time performance. Experimental results show that the system can meet commercial requirements.

3.2 Comparison with SSD

The SSD deep learning network is also a deep learning network using VGG16 as the basic network. The SSD network has advantages in training speed and running speed, but it performs poorly on the detection of small targets [7]. Because the data set required to train the SSD network is also in the VOC2007 format, in order to compare its performance with Faster R-CNN, it uses the same data set to train it. In the case of the same number of iterations, the SSD network has very poor detection results. Although, bread can be detected, the classification of each bread is almost completely wrong. The Bread Recognition System based on Faster R-CNN deep learning network proposed in this paper is superior to the bread recognition system based on SSD deep learning network in the recognition accuracy. The results of testing 100 images in two models are shown in Table 3. The results show that although the SSD model is slightly faster than Faster R-CNN, the accuracy is far less than Faster R-CNN.

Table 3. Comparison of SSD and Faster R-CNN

Model	Test time(s)	Accuracy
SSD	9.15	9%
Faster R-CNN	10.95	100%

4 Conclusion

Deep learning applied to image processing has broad application prospects in the field of new retail and unmanned stores. Based on the deep learning framework of tensorflow1.9.0, this paper constructs a bread recognition system based on Faster R-CNN deep learning network, and trains on the basis of self-built dataset to realize bread recognition of input images. The system accurately identifies each type of bread in the input image and calculates the amount of each bread in the input image. The experimental results on the test set show that the system has good robustness and can still complete the bread recognition task under different lighting conditions and photographing angles. In the future, in order to prevent malicious stacking of bread to reduce the number of breads in the input image, the bread recognition system that combines the depth information with the Faster R-CNN deep learning network will be further developed to further improve the accuracy of recognition. In order to apply the bread identification system to unmanned stores, we will continue to study how to combine the bread identification system with the payment system to improve its application value.

Acknowledgements

The work was supported in part by the Key Laboratory of Digital Fujian on IOT Communication, Architecture and Security Technology under Grant 2010499.

References

[1] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in: Proc. Advances in Neural Information Processing Systems, 2015.

[2] H. Jiang, E. Learned-Miller, Face detection with the faster R-CNN, in: Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017.

[3] Q. Fan, L. Brown, J. Smith. A closer look at Faster R-CNN for vehicle detection, in: Proc. 2016 IEEE Intelligent Vehicles Symposium (IV), 2016.

[4] J. Li, X. Wang, H. Su, Supermarket commodity identification using convolutional neural networks, in: Proc. 2nd International Conference on Cloud Computing and Internet of Things (CCIOT), 2016.

[5] H. Naito, Commodity recognition apparatus and method for recognizing commodity by the same, U.S. Patent 9208383, December, 2015.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single shot multibox detector, in: Proc. European Conference on Computer Vision, 2016.

[7] G. Papandreou, I. Kokkinos, P. Savalle, A modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015.